

Sind Noten nützlich - und nötig?

Ziffernzensuren und ihre Alternativen im empirischen Vergleich

*Eine wissenschaftliche Expertise¹
des Grundschulverbandes
erstellt von der Arbeitsgruppe Primarstufe
an der Universität Siegen*

*von Hans Brügelmann mit
Axel Backhaus, Erika Brinkmann (Gast)
Hendrik Coelen, Thomas Franzkowiak
Simone Knorre, Barbara Müller-Naendrup
Elisabeth Oser, Sara Roth*

*Zusammenfassung für die Presse
Sperrfrist: 14. Juni 2006, 0:00 Uhr*

¹ Der folgende Text ist eine Zusammenfassung der ausführlichen *Expertise*, in dem die einschlägigen Publikationen, insbesondere die empirischen Studien, differenziert ausgewertet und belegt sind. Verweise im Text beziehen sich auf die entsprechenden Kapitel der *Langfassung*. Diese ist erhältlich gegen eine Schutzgebühr von 18 € über den Grundschulverband, Niddastr. 52, 60329 Frankfurt, Tel. (069) 77 60 06 E-Mail: info@grundschulverband.de Unsere Analysen legen grundlegende Probleme einer pädagogischen Leistungsbeurteilung offen, die konkreten Folgerungen beziehen sich aber vor allem auf die Grundschule.

Das Wichtigste auf einen Blick

Ein erstes Problem in der Diskussion ist die unklare Begrifflichkeit. In diesem Gutachten verstehen wir - sofern nicht ausdrücklich etwas anderes gesagt wird - unter »Noten« bzw. »Zensuren« *Ziffernnoten*, die zur *formellen* Beurteilung verwendet werden, z.B. bei Klassenarbeiten oder in Zeugnissen.

»Leistungsbeurteilung« verwenden wir als Obergriff für die *Beschreibung* und die *Bewertung* von Leistungen - zwei unterschiedliche Formen ihrer Rückmeldung, die auch in der Umsetzung sorgfältig zu trennen sind.

Zwei zentrale Erträge unserer Analysen vorweg

▲ Leistungsbeurteilungen haben in unserem Schulsystem nicht nur unterschiedliche, sondern oft widersprüchliche Funktionen zu erfüllen: als Beschreibungen *orientieren* sie über den individuellen Leistungsstand und über Möglichkeiten zu dessen gezielter Verbesserung; sie sind damit ein pädagogisches Medium zur Förderung des Lernens. Als Bewertungen dienen sie der *Disziplinierung* und *Selektion*. Spätestens seit der UN-Kinderrechtskonvention erweist sich ein *hierarchisches* Verständnis von Leistungsbeurteilung als nicht mehr zeitgemäß. Nicht Anpassung und Gehorsam, sondern Mitbestimmung und Selbstverantwortung sind vorrangige Erziehungsziele einer demokratischen Schule. Schärfere Selektion führt im Übrigen nicht zu besseren Leistungen wie die internationalen Leistungsstudien gezeigt haben.

Empfehlung:

Eine demokratische Schule hat die Persönlichkeit der SchülerInnen durch Formen der Dokumentation und der Bewertung von Leistung zu achten, die ihre Selbstständigkeit fördern statt Abhängigkeiten zu verstärken. Einem solchen Verständnis von Schule sind Noten als Belohnungs-/Bestrafungssystem nicht mehr angemessen. Vielmehr ist die Fähigkeit zur Selbsteinschätzung und zum konstruktiven Umgang mit Kritik zu fördern. Hierfür ist eine sachliche Information der SchülerInnen über den individuellen Stand ihrer Lern- und Leistungsentwicklung unerlässlich.

▲ Ziffernnoten sind immer noch die häufigste Form formeller Leistungsbewertung in der Schule. Aber die Forschung zeigt seit langem: Noten sind nicht in der behaupteten Weise für das Lernen nützlich und sie sind erst recht nicht nötig. Sie betonen einseitig die Bewertungsfunktion - können aber auch diese wegen ihrer mangelnden Aussagekraft, Vergleichbarkeit und Objektivität nicht angemessen erfüllen. Es gibt deshalb keinen Grund, auf ihnen zu beharren, zumal sie darüber hinaus etliche unerwünschte Nebenwirkungen haben.

Empfehlung:

Ziffernnoten sind zu ersetzen durch differenziertere Formen der Dokumentation und der Bewertung von Leistungen. Rückmeldung und Bewertung sind klar zu trennen. Beschreibungen sollen den Leistungsstand bezogen auf konkrete Lernziele und die individuelle Entwicklung darstellen. Das lernförderliche Potenzial differenzierter Rückmeldungen wird in der Praxis aber nur dann zur Geltung gebracht werden können, wenn die entsprechenden Rahmenbedingungen geschaffen werden: vor allem durch eine Verringerung des Selektionsdrucks im Bildungssystem und durch eine fachliche Qualifizierung der LehrerInnen.

Die Ergebnisse unserer Analysen im Einzelnen

▲ Noten sind *informationsarm*. Dieselbe Punktzahl in einer Probe kann Ausdruck ganz unterschiedlicher Leistungen sein. Entsprechend werden unterschiedliche Leistungsprofile mit derselben Ziffer belegt.

Empfehlung:

Leistungen sollten nicht nur bewertet, sondern zunächst differenziert beschrieben werden. Die individuellen Stärken und Entwicklungsmöglichkeiten verdienen eine besondere Beachtung. Wo Noten vorgeschrieben bleiben, sind sie schriftlich oder im Gespräch inhaltlich zu kommentieren.

▲ Noten sind *nicht vergleichbar*, da die Bewertung in der Regel auf den Durchschnitt einer Klasse bezogen wird. Je nach Leistungsniveau der einzelnen Klasse wechseln die Noten für dieselbe Leistung. Zudem sind die Maßstäbe je nach Fach und Altersstufe unterschiedlich.

Empfehlung:

Soweit Leistungen überhaupt vergleichend beurteilt werden, sollten *Bewertungen* auf klassenübergreifende Stichproben bezogen werden (z.B. in Form von Prozentrangplätzen in normierten Tests). Allerdings muss bedacht werden, dass Tests nur bestimmte Arten von Leistungen erfassen können. Sie dürfen deshalb nicht zum heimlichen Curriculum werden - zum Beispiel über zentrale Lernstandserhebungen.

▲ Vorrangig orientiert sich die Leistungsbewertung immer noch am Vergleich mit einer Bezugsgruppe. Die Dominanz des sozialen Vergleichs bei der Notengebung *widerspricht* allerdings den *rechtlichen* Vorgaben. Sie hat zudem *negative* Auswirkungen auf die *Lernmotivation* von leistungsschwächeren SchülerInnen, und sie beschädigt die Kraft *intrinsischer Motivation* auch bei den leistungsstärkeren.

Empfehlung:

Die Bewertung von Leistungen muss sich deshalb stärker an Lernzielen und in der Grundschule vor allem am individuellen Lernfortschritt (Entwicklungsnorm) orientieren.

▲ Zensuren sind Urteile von Lehrpersonen. Sie basieren in der Regel auf informellen Leistungsproben und Beobachtungen. Diese Daten und ihre Bewertung in Form von Noten haben sich als *nicht zureichend gültig* (»valide«), *personunabhängig* (»objektiv«) und *verlässlich* (»reliabel«) erwiesen. Soziale und ethnische Herkunft, Geschlecht, aber auch Verhaltensauffälligkeiten und persönliche Sympathie führen zu systematischen Verzerrungen der Beurteilung.

Deren Fehleranfälligkeit verliert erst an Bedeutung, wenn sie nicht zu Selektionszwecken eingesetzt werden. Für lernförderliche Rückmeldungen sind Empathie und eine persönliche Beziehung sogar von Vorteil. Im Übrigen kann die Nutzung von standardisierten Tests zwar die Datengrundlage von Beurteilungen erweitern; ersetzen können die - ebenfalls fehleranfälligen - Tests das Lehrerurteil aber nicht.

Empfehlung:

Leistungen sind möglichst zu mehreren Zeitpunkten und in unterschiedlichen Aufgaben/Situationen zu *erfassen*. Vorstrukturierte Portfolios bieten eine gute Möglichkeit, Leistungen differenzierter, aus verschiedenen Perspektiven und in ihrer Entwicklung über die Zeit hinweg zu dokumentieren. *Bewertet* werden sollten Leistungen möglichst von mehreren Personen, die den Kontext der Leistung und ihrer Entwicklung kennen.

▲ Als Alternative zu Noten werden Verbalbeurteilungen vorgeschlagen. Da sie in der Regel wie Noten auf den Beobachtungen und Bewertungen von LehrerInnen basieren, unterliegen sie aber denselben Einschränkungen, was ihre Validität, Objektivität und Reliabilität angeht. Ihr Vorzug gegenüber Noten: Zumindest vom Anspruch her erfassen sie Leistungen differenzierter, ihre Aussagen lassen individuelle Besonderheiten besser erkennen und sie orientieren sich stärker am Lernfortschritt; darüber hinaus machen sie die Maßstäbe der Lehrperson und die Lernbedingungen deutlicher erkennbar. In der Realität werden Verbalgutachten diesen Anforderungen in vielen Fällen aber nicht gerecht.

Empfehlung:

Um die Vorteile einer verbalen Dokumentation und entwicklungsbezogenen Bewertung von Leistungen stärker zur Geltung zu bringen, sind vier Maßnahmen erforderlich:

- Sensibilisierung von LehrerInnen für die Schwierigkeiten bzw. Fallen von Beurteilungen sowie für die Erwartungen und Lesarten der Zielgruppen;
- eine gezielte Aus-/Fortbildung ihrer Kompetenzen zur Erfassung, Interpretation, Bewertung und differenzierten Darstellung von Leistungen sowie von deren Entwicklung;
- Entwicklung von fachdidaktisch begründeten Kriterien für die Beurteilung von Leistungen - und zwar immer wieder neu in Zusammenarbeit mit den LehrerInnen vor Ort;
- Organisation eines kontinuierlichen kollegialen Austausches über die Maßstäbe und über ihre Anwendung in kritischen Fällen.

▲ Gegen eine ausschließlich verbale Begutachtung unter völligem Verzicht auf Noten haben viele Eltern, LehrerInnen und SchülerInnen immer noch *Vorbehalte*. Diese Skepsis hat aber in der Grundschule und dort besonders bei Personen, die eigene Erfahrungen mit dieser Praxis haben, deutlich

abgenommen. Empirisch *widerlegt* sind die Befürchtungen, Verbalbeurteilungen hätten einen *negativen* Einfluss auf die Leistungsbereitschaft. Bei konsequenter Umsetzung einer ziel- und entwicklungsorientierten Bewertung von Leistungen lassen sich im Gegenteil sogar *positive* Effekte auf das Lernklima in der Klasse sowie auf die Einstellungen und die Motivation der SchülerInnen nachweisen.

Empfehlung:

Wie auch in vielen beruflichen Bereichen sollten zunehmend dialogische Formen einer Verbindung von Selbst- und Fremdbeurteilungen erprobt werden. Die Fähigkeit zur Wahrnehmung und Einschätzung der eigenen Leistung ist gezielt zu entwickeln und in der alltäglichen Anwendung zu unterstützen. Noten »von oben« fördern weder diese Fähigkeit noch die Bereitschaft zur Selbstkritik, sondern provozieren eher Abwehr- oder Ausweichverhalten. Eine symmetrische Beziehung schließt außerdem ein, dass die SchülerInnen nicht nur ihre eigene Leistung, sondern auch die Bedeutung von Lernbedingungen einzuschätzen lernen.

▲ Trotz der durchgängig negativen Befunde über Nutzen und Nebenwirkungen von Ziffernnoten dürfte deren Abschaffung schwierig werden. Dies hängt vor allem mit der frühen und starken *Selektionsorientierung* des deutschen Schulsystems zusammen. Eine rein »technische« Verbesserung des Beurteilungswesens wird deshalb in der Praxis nicht viel bewirken, wenn sich die institutionellen Bedingungen nicht ändern: Verlängerung der gemeinsamen Schulzeit; Abschaffung von Zurückstellungen am Schulanfang, der Wiederholung von Klassen und der Aussonderung in Sonderschulen.

Empfehlung:

Der Förderauftrag der Schule muss bildungspolitisch, in den Schulprogrammen und in der täglichen Arbeit vor Ort Vorrang vor der Selektion gewinnen. PISA-Spitzenreiter wie Schweden - oder im deutschsprachigen Raum: Südtirol - kommen seit vielen Jahren ohne vergleichende Noten aus. Werden dagegen Sanktionen an die Bewertung von Leistungen geknüpft, ist mit einem Rückschlag in dem Bemühen um Verbesserungen zu rechnen. Ranking sowie Selektion *in* und *von* Schulen haben sich vor allem in den angelsächsischen Ländern als pädagogisch kontraproduktiv erwiesen.

Fazit:

Vier Resümees aus vier Perspektiven

Wie bei allen pädagogischen Fragen (und sozialen Phänomenen generell) ist die Befundlage zu Noten nicht auf einen einfachen Nenner zu bringen. Formen der Leistungsbeurteilung wirken unterschiedlich, je nachdem *wie* und in welchem *Kontext* sie eingesetzt werden. Für Folgerungen aus dem Forschungsstand kommt es deshalb darauf an, von welcher Basisannahme man ausgeht: Wer die Beweislast für Veränderungen bei den Reformern sieht, kann zu einer anderen Einschätzung kommen als jemand, der normativ die Förderung des Einzelnen als zentrale Norm und noch uneingelöste Aufgabe der Schule sieht. Vor diesem Hintergrund lassen sich als Ergebnis unserer Analysen vier Folgerungen formulieren:

- ▲ Wer an Ziffernnoten festhalten will, weil sie angeblich objektiv und vergleichbar seien bzw. erforderlich, damit SchülerInnen sich auf die Anstrengungen des Lernens einlassen, findet in der Empirie keine stützenden Belege für seine Position.
- ▲ Auch diejenigen, die Verbalgutachten ablehnen, weil sie negative Auswirkungen auf die Lernbereitschaft und den fachlichen Lernerfolg der SchülerInnen befürchten, können sich auf keine empirischen Daten stützen.
- ▲ Wer andererseits hofft, ohne zusätzliche Maßnahmen, d.h. allein durch die Verordnung von Verbalgutachten Lernbereitschaft und Lernerfolg von SchülerInnen verbessern zu können, wird durch die Befunde zur bisherigen Beurteilungspraxis und ihre Wirkungen ernüchert. Ohne eine pädagogische und didaktische Öffnung des Unterrichts und ohne die Sicherung bestimmter Rahmenbedingungen bleibt eine Veränderung der Bewertung meist erfolglos.
- ▲ Diejenigen aber, die mit dem Verzicht auf Ziffernnoten pädagogische Ziele verfolgen, können mit einer Verbesserung der Unterrichtssituation und der Motivation der SchülerInnen sowie ihres Lernerfolgs rechnen, sofern sie bereit sind,
 - als LehrerInnen sich auf den höheren, aber lohnenden Aufwand einzulassen,
 - als Schulverwaltung die für Evaluation verfügbaren Ressourcen gezielter in die Fortbildung und Unterstützung der LehrerInnen zu investieren und
 - als BildungspolitikerInnen den Selektionsdruck im System zu verringern und Rahmenbedingungen wie die Schüler-Lehrer-Relation zu verbessern.